

Reconstruction of Lamb weather type series back to the 18th century

Gilles Delaygue^{1,2}, Stefan Brönnimann², Philip D. Jones³, Juliette Blanchet¹, Mikhaël Schwander²

1 – Institut des géosciences de l'environnement, Université Grenoble-Alpes, Grenoble, France
(ORCID of GD: 0000-0002-2412-9736 ; JB: 0000-0001-8088-8895)

2 – Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

3 – Climatic Research Unit, University of East Anglia, Norwich, UK (ORCID of PD: 0000-0001-5032-5493)

Abstract The Lamb weather type series is a subjective catalogue of daily atmospheric patterns and flow directions over the British Isles, covering the period 1861-1996. Based on synoptic maps, meteorologists have empirically classified surface pressure patterns over this area, which is a key area for the progression of Atlantic storm tracks towards Europe. We apply this classification to a set of daily pressure series from a few stations from western Europe, in order to reconstruct and to extend this daily weather type series back to 1781. We describe a statistical framework which provides, for each day, the weather types consistent enough with the observed pressure pattern, and their respective probability. Overall, this technique can correctly reconstruct almost 75% of the Lamb daily types, when simplified to the seven main weather types. The weather type series are described and compared to the original series for the winter season only. Since the low frequency variability of synoptic conditions is directly related to the North Atlantic Oscillation (NAO), we derive from the weather type series an NAO index for winter. An interesting feature is a larger multidecadal variability during the 19th century than during the 20th century.

Keywords Lamb weather types; British Isles weather; Climate reconstruction; Little Ice Age, North Atlantic Oscillation

Corresponding author: Gilles Delaygue, gilles.delaygue@univ-grenoble-alpes.fr, ph. +33 476 824 266

1 Introduction

Climatic variability at mid-latitudes is dominated by a seasonally-forced component, and an approximately weekly (synoptic) component. The determinism, trend, and predictability of the latter are still under debate, especially the fact that interactions between the anthropogenic forcing and the natural modes of variability are not yet clear at the regional level (e.g., Corti et al., 1999; Yiou et al., 2007; Brandstator & Selten, 2009). Studies addressing these questions from observations, especially from pressure measurements related to atmospheric dynamics, need datasets consistent over long periods of time (decades to centuries) and over large spatial scales (thousands of kilometres): this has led to the need to homogenise measurements series, and to spatially analyse them. Historically, this analysis started subjectively from looking at synoptic charts. Later, atmospheric models have been used to assimilate these sparse series. By applying the same model over long periods of time, ‘reanalyses’ have brought a fantastic wealth of information about the past. However, reanalyses have some limitations. Their application is limited by the existence and spatial coverage of measurement series. Tests conducted for the Twentieth Century Reanalysis (20CR) as far back as the middle of the 19th century are impressive (20CRv2c products covering the period 1851-2012; Compo et al., 2011), but are still very exploratory. Assimilating sparse observations and starting from a more uncertain initial state allows the model internal variability and its physical biases to contribute more to the analysis. Hence the reliability of reanalysis products sharply decreases back in time (e.g., Ferguson & Villarini, 2013; Krueger et al., 2013), a limitation also seen through the divergence of ensemble simulations (e.g., Brönnimann et al., 2013).

Over the past centuries, gridded datasets exist back to 1850, with a monthly resolution on a global or hemispheric scale (Jones et al. 1987), and with a daily resolution for the north Atlantic region (Ansell et al. 2006). Backward in time, very sparse series exist with a daily resolution, especially for pressure. We propose to analyse these as weather types (WTs), i.e. typical and recurrent meteorological patterns at the scale of hundreds of kilometers. We build on subjective analysis of the synoptic conditions defined decades ago by trained meteorologists with their ‘expert

judgment'. There are two long and continuous time series of daily weather types (WTs) for Europe: the German Großwetterlagen (GWL) series covering Europe over the period 1881-1976 (Hess & Brezowsky, 1977), and the Lamb weather type (LWT) series covering the British Isles over the period 1861-1996 (Lamb, 1972). These WTs were analysed from synoptic charts, essentially surface pressure patterns and winds, and later augmented with upper atmosphere observations. Previous studies have shown that these WTs constitute valuable indices of weather conditions (e.g., Sweeney & O'Hare, 1992; Jones et al. 1993, for the British Isles), of extreme conditions (such as flooding and gale frequency: Wilby & Quinn, 2013; Jones et al., 2014), and of teleconnections (Wilby, 1993; Wilby et al., 1997). These long series have also been used to explore climate change in the European area (e.g., Werner et al., 2000; Werner & Gerstengarbe, 2011). These WT series have almost been forgotten because they are considered too much subjective, unreliable, and less useful than reanalysis, although objective definition of these WTs has been devised and applied to reanalysis (e.g., Jones et al., 2013). Our aim is to apply this WT analysis to the scarce and less reliable early meteorological measurements in order to extend the Lamb WT series back to the end of the 18th century.

The LWT series (usually called the "Lamb weather type catalogue") has been built by manually classifying the daily synoptic conditions prevailing over the British Isles (an area ranging from 10°W to 2°E, and 50°N to 60°N), over the period 1861-1996 (Lamb, 1972). This immense enterprise directed by Hubert Lamb consisted in analysing by eye sub-daily synoptic charts and their succession in time, over almost 50,000 days, in order to characterize the pattern of atmospheric flow (Table 1; Figure 1). Expert judgement is an efficient way to analyse charts of potentially diverse quality. The accuracy of this subjective analysis could be assessed by applying an automatic classification to gridded datasets and comparing this 'objective' LWT series with the original one (Jones et al., 1993; 2013; 2014). In addition to characterising the regional synoptic conditions, LWTs are valuable for parts of continental Europe. On their way to western Europe, the Atlantic storm tracks point towards the British Isles during the whole year (e.g., Woollings, 2010). This

makes the British Isles an ideal location to detect the direction of Atlantic storms towards western Europe, mostly responsible for meteorological variability. This direction is especially sensitive to the north-south gradient of pressure across the Atlantic, whose variability is partly described by the North Atlantic Oscillation (NAO; Wilby et al., 1997).

We extend the LWT series back to 1781, by applying an automatic classification technique to the few continuous daily pressure series available for this period over western Europe. Starting in 1781 bridges a short and mostly unknown part of the Lamb catalogue to the main catalogue. This part mainly covers the years 1781-1786, with a few additional months, for which data exist from the Palatine meteorological society (Kington, 1988). Lamb and co-workers initially envisaged to start their catalogue of daily WTs at the end of the Little Ice Age (Lamb & Johnson, 1966; Kington, 1970), to explore the climate of this period. There were enough data to start with the 1780s decade, known as “peculiar” with frequent cases of weather extremes over the British Isles, related to blocking conditions (Kington, 1988). The lack of time and data prevented Lamb and co-workers from completing this endeavour. Starting in 1781 has two objectives for us: the possibility to test our reconstruction, calibrated on a more recent period, with the earliest part of the catalogue, and to draw the attention of researchers to this existing window into the climate of the Little Ice Age.

Section 2 presents the surface pressure series. Section 3 presents the simplification of the LWT series used in our study and additionally considers more objective weather typing for the British Isles region. Section 4 presents the methodology used to reconstruct the daily WT series. Section 5 presents and evaluates the reconstructions of LWT for winter. Section 6 gives an application of these reconstructions with a North Atlantic Oscillation index.

2 Data sets of pressure records

Meteorological records that are continuous back to the 18th century exist in very limited numbers. Over periods of centuries, different measurement systems have been used, in different conditions (e.g., outside or inside), their location may have moved, so that the measurements must be corrected for different technical biases, and must be made consistent over time (homogenisation). This is

especially critical for pressure measurements which are sensitive to temperature. These corrections require to precisely know the measurements conditions, and this information (collectively referred to as metadata) has rarely survived, so that researchers are left with strong assumptions and it is necessary to compare with nearby stations (if any). Hence, the quality of a dataset remains in itself a challenge, and is even difficult to fully quantify. Generally, the uncertainty increases backward in time, but it strongly depends on the available metadata (see, e.g., Brugnara et al., 2015, for a discussion of several series in Europe).

We chose to use only surface pressure, although other meteorological parameters are available, especially air temperature. Previous studies (e.g., Perry, 1969) have shown that the atmospheric flow direction is directly related to the pressure field and has a less systematic fingerprint in temperature (and there are good physical reasons for that). In fact, we did some preliminary tests using pressure, pressure tendency, and temperature, and the best scores were obtained by using only pressure.

Most of the pressure series were reduced to sea level, except the one from Hohenpeißenberg (977 m): for consistency with the other series we chose to apply this correction, although this has no effect on the scores of the reconstruction. For this correction, we use the daily mean temperature available for this station, and assume a linear decrease of temperature with height (e.g., Eq. 8 of Brugnara et al., 2015). We note that such a correction may not be physically exact for days with temperature inversions, typically in winter, but other series have been corrected with such ‘virtual’ temperature (Moberg et al., 2002).

A strong constraint we chose is to keep the same set of pressure records for the whole period of reconstruction (1781 to 1996). Adding other records when they become available brings more information to the pressure field and should improve the reconstruction of WTs, but this has the drawback of introducing heterogeneities (breakpoints) in the reconstructed series.

In fact, only seven daily records of pressure with very limited numbers of missing days and homogenised over western Europe, could be found for this period (Fig. 1 & Table 2). By using so

few records, our study can be considered a ‘worst case’ test: clearly the working conditions could be improved, by starting the reconstruction later and adding other records. For instance, the pressure record of Armagh, in UK, starts at the very end of the 18th century.

We first want to evaluate the potential of this limited dataset to constrain the pressure pattern at the European scale (synoptic scale). For this, we consider the covariance of the pressure series between the seven stations, calculated separately for each month in order to remove the dominant correlation due to the seasonal cycle. The correlation is then dominated by the daily synoptic variability at the scale of thousands of kilometres. On average Stockholm and Uppsala series have more than 99% of their variability in common. London, Paris, and Amsterdam series have about 80% of their variability in common. Hohenpeißenberg and Paris have about 70% of their variability in common. Milan series is left apart with about 60% of common variability with Hohenpeißenberg. Although both locations are not very far apart, Milan is located south of the Alps in the Mediterranean area, and this may explain this behaviour. The very high correlation level between the Stockholm and Uppsala series is readily explained by their proximity. Additionally, the Uppsala series has been homogenised with respect to the Stockholm series: this affects the mean value of the pressure over some periods, but not the day-to-day variability we are looking at. Hence, both series have been kept for our reconstruction with the aim of smoothing out possible errors at the daily scale.

These covariance values show that the seven series are not so much independent. In fact, the first three eigenvalues of the pressure covariance matrix explain 92% of the January variability across the series. We conclude from these results that such a dataset is unlikely to correctly discriminate the 26 different WTs of the Lamb catalogue (Table 1), and that we have to reduce the number of WTs considered for our reconstruction. This reduction effectively improves the score metrics used to quantify the efficacy of our technique (see below). Such a reduction of the number of WTs has been a common practice in studies using the LWT, by using the seven main types (e.g., by Lamb 1972) or even a unique, composite index like the ‘Progressive-Southerly-Cyclonicity-

Meridionalities' index of Murray & Lewis (1966) used in many subsequent studies. Such reduction has been found to lead to stronger relationships between WTs and meteorological conditions (e.g., Wilby, 1995).

3 Seven weather types to describe the pressure pattern

An obvious simplification of the 26 LWT is to use the 'seven main types' originally used by Hubert Lamb (Lamb, 1950), and considered in studies using the LWT. These seven main types correspond to the five main wind flow directions prevailing over the British Isles, that is, the four cardinal directional types (North-N, East-E, South-W, West-W) and the North-West (NW) directional type considered by Lamb as important, plus the two synoptic cyclonic (C) and anticyclonic (A) types (Table 1). Disregarding hybrid types has been applied in several studies (e.g., Wilby, 1995; Jones et al., 1993). Information from hybrid synoptic types is partly kept by attributing these days to the corresponding 'pure' synoptic type (e.g., type AS counted as the main type A). They could also be attributed to the corresponding 'pure' directional types (e.g., type AS counted as the main type S), but this scheme was found slightly less skilful (see below). For hybrid directional types, such an attribution (e.g., SE to either S or E) is not possible without interpreting the synoptic situation. Hence, the days corresponding to the directional hybrids NE, SE, and SW, were simply unclassified: this eliminates 5.6% of the classified days (which add up to the 3.9% of unclassified days in the catalogue, to give the total of 9.5% in Table 1).

In addition to the original Lamb catalogue, we also use here the LWT catalogue produced by Jones et al. (2013), based on the Jenkinson & Collison (1977) technique of classification. Briefly, this technique is an automatic or 'objective' classification of gridded surface pressure data: it calculates daily parameters related to the vorticity and direction of the atmospheric flow, used to infer the corresponding LWT. Jones et al. (2013) applied this technique to the 20CR and National Center for Environmental Prediction (NCEP) reanalysis to produce a daily series of LWTs over the 1871-2010 period. This series is operationally updated using NCEP (<https://crudata.uea.ac.uk/cru/data/lwt>). The series is close to the original Lamb subjective catalogue,

with some differences in the WT frequencies and in decadal trends such as the absence of a long-term decline in the westerly type (Jones et al. 1993; 2013; 2014). Instead of simplifying the 26 WTs of this series as with the original Lamb catalogue, we adapt the algorithm of Jones et al. (2013; their Appendix A1), and apply it to the daily flow parameters they have published (Westerly flow W, Southerly flow S, and vorticity Z components). Hybrid synoptic types are also attributed to the corresponding ‘pure’ synoptic type. Hybrid directional types are not considered, and because the flow direction is precisely known (W & S parameters), every day corresponding to such a hybrid type can be attributed to one of the five main directional types. For instance if the flow direction is 140° clockwise, the hybrid type would be SE, and because this direction is closer to S than to E, the day is classified as the main type S. This saves information from a large number of days since only 2% of the days are left unclassified when the flow condition is uncertain (Table 1). An additional LWT catalogue is hence built over the period 1871-1996. For clarity, in the following we will refer to it as the objective catalogue, and to the original LWT as the ‘subjective’ catalogue.

4 Classification of daily pressure pattern as LWT

We apply an automatic classification of the daily SLP anomaly values at the seven stations to produce series of LWT. This classification is based on combining the footprints of each LWT at each station. We call these ‘centroids’. For each station, the centroid of each LWT is the averaged SLP anomaly associated to this particular LWT (i.e., calculated over the days classified with this LWT in one catalogue). The SLP series have been normalised separately for each month in order to remove 1. the difference in variability between stations and 2. the seasonal cycle of pressure. Centroids are thus calculated separately for each month. The calibration period over which the centroids have been calculated is 1861-1996 with the subjective catalogue and 1871-1996 with the objective catalogue. Shorter periods have been tested as well, to test the classification efficiency.

To illustrate this technique, Figure 1 shows the synoptic situation (pressure pattern) of one particular day. This day is classified in the Lamb catalogue as westerly, one of the most frequent types (Table 1): a low pressure system is located between Iceland and the British Isles, whereas a

high pressure crest stretches from the Azores towards Europe. Figure 2 shows the monthly centroids for each station: pressure is below (above) average at all stations for a cyclonic (anticyclonic) situation, and pressure anomalies are contrasted for the other WTs. For instance, in the synoptic situation shown by Figure 1, pressure is above average at the Hohenpeißenberg and Milan stations (this points to either types W, S, or A), and below average at Stockholm (and Uppsala), which identifies type W.

By using the LWT catalogues to classify days and their SLP anomaly, the SLP variability is clustered among the centroids. The efficiency of these clusters to statistically describe the SLP variability can be quantified with the proportion of SLP variance explained by these clusters, the so-called explained variance (EV): EV is the proportion of intertype (inter-WTs) variance in the total SLP variance (Equation A1 in the Appendix). A higher EV allows a better discrimination of LWTs by SLP. With the subjective catalogue used to define the centroids, EV amounts to 37%. With the objective catalogue used to define the centroids, EV amounts to 35%. This slightly lower explained proportion of variance (slightly higher proportion of intratype variance) may originate from the days classified as hybrid directional types in the original catalogue (Section 3). If calculated over the shorter calibration period 1945-1996, EV does not change significantly (less than 1%).

The technique of classification used here is quite standard; it is essentially the same as the one used by Schwander et al. (2017) to reconstruct WTs over Central Europe. The main difference is that we follow a probabilistic approach: instead of selecting the most likely (probable) WT for each day, we consider all WTs consistent with the observed pressure pattern and their associated probability. Only the methodology is presented here, the whole process (flow chart), technique and equations are detailed in the Appendix.

We use the Mahalanobis distance (D), a standard metric which accounts for the pressure covariance between stations (e.g., Wilks, 2011), to calculate the seven distance values D_i between the observed daily pressure values and the i centroids (Equation A2 in the Appendix). However, it may happen that a synoptic pressure pattern does not correspond to any LWT and cannot be

classified ('unclassified type'). This case may also arise from measurement bias or correction which makes the set of pressure values inconsistent with the LWTs (independently of the actual synoptic pattern). Thus, we check each day whether the set of pressure values correctly projects onto the centroids. For this we test whether the shortest distance D_i is acceptable with respect to the distribution of distances around the centroids (the test is fully described in Section 6 of the Appendix). If not, this day is not classified. About 3% of the days with pressure values were rejected by this test. (This proportion is not very sensitive to the risk of accepting a wrong rejection.)

Based on these distances D_i , the procedure used in most studies is to classify each day with the closest weather type i , since the shortest distance D_i maximises the probability of the WT i (Equation A3 in the Appendix). However, the choice of WT with the highest probability does not make it necessarily correct: if the second maximum is not very different from the highest, there is a chance that both corresponding WTs could match the actual synoptic pattern. A reason for this is the large overlap of pressure distributions around centroids at each station, as suggested by Figure 2. The limited dataset of pressure and the locations of the available stations may under-constrain the classification of pressure patterns. With no possibility to really improve this situation, we propose to get around this limitation by calculating the probabilities P_i that each of the seven WTs i matches the observed pressure pattern (Equation A4 in the Appendix). These probabilities give information on how well the WT is constrained each day, and this will help decide how much trust can be put on the reconstructed WTs.

5 Reconstructed LWT series and their evaluation

5.1 Results of the reconstruction

The LWT series over the period 1781-1996, reconstructed from the subjective catalogue, are shown in Figure 3 for winter (DJF), as the season with the strongest atmospheric dynamics and with the strongest climatic impacts of synoptic variability. The proportions of WTs are difficult to discuss separately, and for this reason previous studies have used their combination to form indices of

circulation; this will be done later in this study. Some features are still worth noting. There is a strong multidecadal to centennial variability, as shown by the historical high frequency of Easterlies over the period 1840-1860, approximately, and the historical high frequency of Westerlies over the period 1880-1920, approximately. Westerlies frequency can be compared with the extension of GroßWetterLage W series back to 1780 published by Beck et al. (2007) (their Figure 9, for January), and also with the western index based on wind direction derived by Barriopedro et al. (2014). All three series agree on low frequencies of Westerlies from the end of the 18th century to approximately 1850. However, the 1850-1870 period of minimum frequencies of GWL W is not present in the two other series. The broad maximum found in our series over the period 1880-1920 is consistent with the maximum found in the two other series, although it started almost 10 years earlier. The comparison can also be made for the frequencies of Southerly types: both series of GWL S reconstructed by Beck et al. (their Fig. 9) and our Southerly series (Fig. 3) started with historical low values and display a consistent increasing trend towards the 20th century. However, the GWL S series kept increasing, whereas our series shows a maximum over the 1910-1930 and decreased afterwards.

The end of the 18th century features high frequencies of Northerly and Easterly types. These types are usually associated with blocking conditions, and these high frequencies are in line with the observations of “peculiar meteorological conditions” for this period (Kington, 1988). However, the turn of the 19th century is obscured by frequent unclassified days, and this largely impedes analysing the earliest part of our series. Days are unclassified for two reasons: (1) if (at least) one pressure measurement is missing among the stations (this case corresponds to 55% of the unclassified days over the period 1781-1996), or (2) if the observed pressure pattern is considered too distant from the centroids to match any weather type (Section 4; 45% of unclassified days). As is apparent in Figure 3, missing data are mostly grouped over few winters in a row, for which archives and data are missing. Days which are unclassified because of a non-matching pressure pattern are more regularly distributed. The DJF frequency of this latter type of unclassified days

changes over three periods: between 1781 and approximately 1840 it remains below 10% with an average value of ~5%, between approximately 1840 and 1870 it decreases below 5%, after 1870 it decreases below 2-3%. This increase in the frequency of unclassified days backward in time is due to an increase of the distance between the pressure pattern and the centroids. Such increase may be due to an uncorrect homogenisation of pressure series, since the data required for a precise homogenisation (e.g., temperature, type and location of barometer) are usually missing in the past.

5.2 Evaluation of the reconstruction

A first evaluation of the reconstructed WT series is the comparison with the catalogue used for the calibration. As an indicator of the general synoptic activity, the annually averaged frequencies of WTs are compared in Table 3 for the reconstructed and the calibration periods. In general, the most frequent types (Anticyclonic, Cyclonic, and Westerly) are correctly detected by the method, but with some bias, up to 9% (for the Anticyclonic type). Biases are in the same direction (i.e, they have the same sign) when using the subjective or objective catalogues (except for the Anticyclonic type). This suggests that part of these biases are related to the available set of pressure records, that is, the capacity of these records to detect the synoptic pattern. The last columns of Table 3 display the frequencies of WTs restricted to the days correctly classified: frequencies are much closer to the ones of the original catalogues, except for the dominant Cyclonic & Anticyclonic (A&C) types, still biased high by 5-6%. This means that our classification technique too frequently attributes the pressure patterns to the directional types (higher proportions in the total series than in the correct days) at the expense of the A&C synoptic types. Figure 4 compares the variations of WT frequencies for winter between the reconstructed series (Fig. 3) and the catalogues used to build them. Overall their correlation is high (up to $r=0.85$), and is slightly better when using the objective catalogue. There is no systematic degradation in the past of the reconstructed series (i.e., no increase of the difference with the reference catalogue). For instance the bias in Easterly proportions does increase in the past before ca. 1950, but conversely the bias in Anticyclonic proportions is stronger in the latter half of the 20th century, and the bias in Cyclonic proportions is

almost stable. The bias in the easterly WT is remarkable in a relative sense, although it concerns few days; it is still important because easterly WT is typical of non zonal conditions. This bias seems to be due to our classification method, which attributes too frequently unclassified and southerly synoptic conditions to easterly WT (not shown).

We further compare the reconstructed series with the calibration catalogue at the daily level. A simple metric is the proportion of days correctly classified, usually called the matching score. The matching scores are 56% with the subjective catalogue and 59% with the objective catalogue (Table 4). We can test the sensitivity of these values to the calibration period: with centroids calibrated over the shorter period 1945-1996 (instead of 1861-1996), these matching scores slightly decrease to 55% and 57%, respectively. Calculating the matching scores over the 1861-1944 period (i.e., not included in the 1945-1996 calibration period) further reduces the matching scores to 52% and 54%. These score values have been calculated with the most likely WT for each day. We can calculate by how much a score is improved when considering also the second most likely WT. Assuming that a day is well classified if either of the two reconstructed WTs matches the WT of the calibration catalogue, the matching scores increase to 75% and 78%, respectively. This increase by 20% of the score is important. It can be related to the consistency of both WTs (when they both exist, that is, in 91% of days) in terms of synoptic conditions. Especially, when both WTs are directional types (N, S, etc.; in 15306 days), they are adjacent in 97% of these days (e.g., NW & N, but not N & S). Also, when one likely WT is a 'pure' synoptic type (A or C), the other likely type is the reverse 'pure' synoptic type (C or A) in less than 0.6% of these days. This proximity of both WTs, and the high score value obtained when considering the second likely WT as an alternative, motivate the consideration of several WTs for each day rather than only the most likely one (as usually done). We will later use a complementary metric to show that the score increase of 20% is significant and that considering the second likely WT as an alternative actually improves the skill of our classification.

Our reconstructed series can be further compared with days classified over the years 1781-

1786, analysed by Lamb and co-workers in addition to the period 1861-1996 forming the main catalogue (Lamb, 1994). Other very short periods of special interest, before 1861, were also classified (Kington, 1988). These data have been digitised to complement the main catalogue. Over a total of 1951 days, the matching score of our reconstructed series is 46% and 64% considering either one or two WTs each day. This score is lower than over the main catalogue (56% and 75%, respectively; see Table 4) but the comparison is limited for different reasons: the number of tested days is much lower (1951 vs. 78892); the pressure data are more likely to be biased; and the proportions of WTs are different especially because these periods were selected for their extreme meteorological conditions by Lamb and colleagues (Lamb 1994).

As cited above, a long series of pressure does exist at the northwestern station of Armagh (Northern Ireland). Although we chose not to use it (because our reconstruction would start later), the question arose of the improvement of our (shorter) reconstruction by including this series: the matching score only increases by 2%. This very limited improvement may seem surprising; however, the pressure centroids at Armagh mostly discriminate the Easterly and Northerly patterns, which are rare types (about 8% of the days in total).

The matching score is complemented with another metric, the skill score S , which assesses the skill of our reconstruction technique to ‘do better’ than by chance. Chance means here picking randomly one WT each day based on the climatological (average) frequencies of the WTs (Equation A5 in the Appendix). We use this metric to test whether considering two likely WTs for one day, instead of the most likely one, not only improves the matching score (which is trivial) but also the skill score. This is in fact the case (Table 4), and this underlines the added-value of our probabilistic approach. (Note that the values in the last row of Table 4 have been calculated over a much smaller number of days, with WT proportions different than over the full catalogue: for these reasons they can be only tentatively compared to the other S values.)

A last metric, the probability of the classified WTs (Equation A4 in the Appendix), is used to evaluate how well our reconstruction methodology can discriminate among WTs. A high probability

for a WT means that this WT has a higher probability than the other WTs to match the observed pressure field (however this does not imply that this WT is the correct one). The probability of the most likely WT has a median value of about 51%, a mean value of 54% ($\sigma = 17\%$), and a high density of values close to 100%. Both subjective and objective reconstructed series of WTs have very similar distributions of these probabilities. This median (or average) probability is not very high: this means that our classification technique is not very efficient at discriminating the WTs, a limitation probably due to the distribution of the stations and their limited number. However, for days with two likely WTs (more than 90% of the days), the ratio of both probabilities has a median value of about 2, and their sum has a median value of about 80%. These figures show that when two WTs have been classified, the more likely WT is twice more probable than the second most likely WT, and the probability that other WTs match the observed pressure field is quite low (a fraction of 20%, for instance the mean value of the third most likely WT is about 12%). This underlines the interest of considering an alternative WT for each day (and the probabilities of the WTs) instead of only the most likely WT.

6 Reconstruction of a North Atlantic Oscillation index

The so-called NAO is the first mode of pressure variability in the north Atlantic region (e.g., Wanner et al. 2001). It is a swing in the meridional gradient of pressure, particularly marked in winter. This pressure gradient is related with the direction and speed of storm tracks across the Atlantic, which are both major sources of synoptic variability over western Europe. WTs like the Großwetterlagen and the Lamb WTs, have been defined and used by meteorologists to analyse this variability (e.g., Lamb 1950). The frequencies of these WTs have been used to describe low frequency modes of circulation, from the zonal circulation (in the 1950s) to the NAO later (e.g. Murray & Lewis, 1966; Wilby et al., 1997, using LWT; Michelangeli et al., 1995). These studies showed that the synthetic description of the synoptic activity by WTs is adequate to analyse its low frequency modes of variability.

We build on these previous studies and show that the Lamb WT frequencies can be used to

build and extend in time an NAO index. Murray & Lewis (1966) proposed the ‘Progressive index’ to quantify the progression of storm tracks moving eastward across the British Isles. This index was calculated as a weighted sum of Lamb WT frequencies. Weights were chosen to account for the opposing contributions of WTs to this progression. This Progressive index correlates quite well with standard station-based and Principal Component (PC)-based NAO indices (typically $r = 0.73$ to 0.75 , ± 0.07 at 95%). We extend their study by determining weights from multiple regressions with NAO indices, that is, weights which maximise the correlation between WT frequencies and an NAO index. We use the Azores-Iceland station-based NAO index and the PC-based NAO index of Hurrell (1995), available over the periods 1864-2014 and 1899-2015, and the Iberia-Iceland station-based NAO index revised by Vinther et al. (2003), starting in 1823. The PC-based index accounts for spatial changes over time of the NAO pressure centres, contrary to a station-based index, and should better describe the NAO variability (e.g., Hurrell, 1995; Osborn et al., 1999). However, it covers a shorter period.

We first calculate an index based on the extended winter (DJFM) WT frequencies of the subjective Lamb catalogue over the period 1861-1996 (plus the days classified over the years 1781-1786). The multiple regression between the seven WT frequencies (number of days over DJFM) and this PC-based NAO index, over the overlap period 1899-1996, provides us loadings used as weighting factors to calculate the index shown on Figure 5a. The correlation of this index with the PC-based NAO index is $r = 0.81$ (0.74 - 0.86 at 95%). These loadings are somehow different than the weights used for the Progressive index of Murray & Lewis (1966), but close in terms of contributions of the different WTs to the indices. This is consistent with the fact that the positive phase of the NAO corresponds to a zonal synoptic circulation. Unclassified days are difficult to account for: in calculating winter (DJFM) occurrence of WTs, unclassified days are not considered, so that our index can be evaluated for every winter. However, with a calculation done with fewer days, the result may be biased. Our index is especially sensitive to such bias because its calculation puts weights on the infrequent northerly and easterly types (Table 3), which may be

disproportionately affected by missing days. Hence, we use a conservative maximum of 20% for the number of unclassified days per winter (i.e., 24 days over the 121 days of DJFM), above which the index value is not considered (missing values on Figs 5a and 5b).

Our index captures well the decadal variability of the PC-based NAO index (Fig. 5a). We find values consistently low during the 1870s, 1890s, and the 1950s-1960s; and consistently high values during the 1900s and the 1990s, although we miss a few winters of the latter because of too many unclassified days. Our index also captures well the interannual variability of this PC-based NAO index. The most extreme winters are captured, although not with the same absolute values. For instance, the 1968-69 and 1988-89 winter record values are not as low with our index, and conversely the 1995/6 winter value of our index is more extreme than in the PC-based NAO index. The lowest value is found for the winter 1783/4, and the value for the following winter 1784/5 is the third lowest just above the one for 1995/6. Such low values for these winters 1783/4 and 1784/5 have been found in other reconstructions of NAO variability based on SLP (Cornes et al. 2013) or on proxies (D'Arrigo et al., 2011; Ortega et al., 2015). We note that these winters directly followed the strong volcanic eruption of the Laki in 1783.

Other regressions were tested, using either the subjective or the objective WT catalogue, and the standard NAO indices cited above. The regression loadings and the correlation level are robust among these tests, with a typical correlation level of $r = 0.8$. The highest correlation is $r = 0.86$ (0.80-0.90 at 95%) over the period 1899-2015; it is found between the Vinther et al. NAO index and the index regressed with the objective WT catalogue. For comparison, the correlation level between the station-based and PC-based NAO indices is typically $r = 0.84$ to 0.88 for winter (Osborn et al., 1999). This suggests that our WT-based index is as good as any other one to describe the NAO variability.

Last, we calculate an index over the period 1781-1996 based on the reconstructed Lamb (subjective) WT series described above, regressed to the PC-based NAO index of Hurrell (1995). For the regression, instead of using the most likely WT each day, an alternative WT with its

probability is considered in order to estimate the uncertainty on the WT frequencies. For this, the reconstructed series was resampled 1000 times. Each day with an alternative WT, a number following a uniform distribution over $[0;1]$ is randomly generated: if it is lower (higher) than the highest probability, the most likely (the second most likely) WT is chosen. Each daily sample of WTs was averaged over winters (DJFM) to produce a series of winter WT proportions, regressed against the DJFM PC-based NAO index over the 1899-1996 period. The loadings of this regression were applied to the WT proportions over the full 1781-1996 period to produce an NAO index time series. From these 1000 time series, we can calculate the average and standard deviation to estimate its uncertainty (0.25, constant over the whole series). The correlation between the averaged WT-based index and the PC-based NAO index is $r = 0.82$ (0.74-0.87 at 95%) over 1899-1996. Figure 5b compares this index to the station-based (Iberia-Reykjavik) NAO index extended by Vinther et al. (2003), which is the longest available index based on SLP series (starts in the winter 1822/3). Missing values correspond to winters with a high occurrence of unclassified days (Fig. 3), and this concerns many winters at the end of the 18th century (including the 1783/4 one, with the lowest value in Fig. 5a, but not 1784/5 with the lowest value in Fig. 5b). Our index shows consistently low values over the 1840s-1870s period (approximately 35-40 years), with among the lowest values of our index (8 of the 21 lowest values over 1781-1996, especially the 1844/5, 1864/5, and 1869/70 winters), with the exception of few years around 1850 with a positive index. This makes this period much more consistently low than the 1950s-1960s period. These low values happened after a long decreasing trend starting from the 1810s-1820s decades with consistently high values. Such decadal persistence of low values is not so obvious in station-based NAO indices like the Vinther et al. one, which displays a stronger variability at the interannual scale. Still, the correlation with this Vinther et al. NAO index over their whole common period (1823-1996) is $r = 0.82$ (0.77-0.87 at 95%). Note that using our reconstructed WT series based on the objective catalogue leads to a very similar index (within the 0.25 uncertainty given above), but with more missing winters at the beginning of the index. This is why we discuss and show on Figure 5b the index based on the subjective

reconstructed WT series.

Our index can be compared to circulation indices partly related to NAO variability. Especially, a ‘western index’ of the zonal flow over western Europe has been derived from the SLP series of London and Paris (Slonosky et al., 2000; Jones et al., 2003; Cornes et al. 2013), and from wind directions in the English Channel (Barriopedro et al. 2014). Cornes et al. (2013) did an extensive work to show that their index mostly capture extreme negative NAO conditions. Although our index is not independent from this western index since they both use the same SLP series, a comparison of it with their DJFM index (their Figure 4g) shows some consistency for negative values, for individual winters (especially 1784/5, 1916/17, 1995/96), and for some decades (1850-60, 1935-50). However, the consistently negative period of 1855-1890 in our index is not so obvious, neither in the index of Cornes et al. nor in the index of Barriopedro et al. Early periods of low values common to the three indices can still be found, especially the end of the 18th century (1780-1790), and the decrease over 1820-1830.

Other long series of NAO-like index have been derived from proxies (mostly tree rings), but the comparison is even less direct, as can be shown by the low correlation level with our series (for instance $r = 0.21$ over the 1781-1969 period with the index of Ortega et al. 2015).

At the multidecadal scale, our index has a marked decreasing trend from the 1820s to the 1870s, followed by a rapid increasing trend into the beginning of the 20th century. By contrast, the 20th century is marked by lower multidecadal variability but with larger interannual variability, culminating at the end of this century (and, in other indices, at the beginning of the 21st century). These features contrast with the conclusions of Goodkin et al. (2008) from an NAO-like index based on a filtered record of Sr/Ca from a Bermuda coral: their index has a stronger multidecadal variability during the 20th century than during the 19th century. We note, however, that the coherence between their Sr/Ca record and the station-based NAO index of Hurrell (1995) is quite low (squared coherence below 0.25, see their Figure 2a), and that the bias on this Sr/Ca record (due to missing annual bands) should be maximum over the 1830-1865 period (which is precisely the

period of maximum multidecadal variability in our index).

A relevant question is whether there is any interest in calculating an NAO index from WT series rather than directly from the SLP series used to infer these WTs. The earliest efforts to define an NAO index in the 1920s used in fact weighted averages of SLP and temperature series, an approach not far from a principal component analysis of the series (Wallace, 2000). At least two arguments for using WTs can be given. First, an index can be effectively calculated by applying the same technique of multiple regression to the pressure series used in this study (Table 2). If regressed to the PC-based NAO index of Hurrell, the correlation between this generated index and the PC-based NAO index is $r = 0.84$ (for DJFM over 1899-1996), compared to 0.82 with the WT-based index. This means that the same proportion of NAO variability can be described with the SLP series and the WT series. However different parts of this variability are described: our index based on WT-series has a larger decadal variability and fewer extreme values compared to the station-based NAO index of Vinther et al., but less multidecadal variability than the PC-based NAO index (which has a red spectrum). A second argument for using WTs to generate an NAO index is that WTs correspond to real synoptic conditions. In addition to the SLPs used in this study, WTs can be constrained by other meteorological observations like wind directions, temperature, etc. Conversely, WTs can be used as predictors of meteorological conditions, like hydrological indices (e.g., Wilby & Quinn, 2013) and NAO extremes. For instance, the very low negative value of the 1844/5 winter (NAO index of -2.0) is due to much more frequent Easterly and Northerly types, and less Westerly types, whereas the extremely positive value of the 1988/9 winter (NAO index of +2.5) is due to much more frequent Westerly types, and almost no Easterly and Northerly types.

7 Conclusions

We have set up a statistical framework in order to classify SLP patterns as daily Lamb WTs, and to estimate their probabilities. Such a probability estimate is generally missing in statistical techniques of classification, because only the most likely WT is considered. This deterministic approach is only valid if the associated probability is high enough. This is not the case in our study,

because the classification technique is applied to few long series of SLP, which due to scarce locations, and probable heterogeneity lead to rather low probabilities for WTs. To cope with this problem, we test whether an alternative WT exists to the most likely WT, consistent with the SLP pattern, and if so calculate its probability. Accounting for this alternative WT improves the skill of the classification technique, and greatly improves the number of days correctly classified (typically from 55% to 75%, compared to 18% and 36% by chance). This probabilistic approach should be useful in other contexts, and calculations given in the Appendix allow one to estimate the probability of the most likely WTs.

Climatic variability in winter is well explained by synoptic conditions in western Europe, so that WTs are useful to describe this variability and its modes, especially the NAO. Our reconstructed series of Lamb WTs show weak zonal flow around the turn of the 18th century, consistently with other indices. Further, the frequencies of Lamb WTs have been regressed to a PC-based NAO index, with a very high correlation. The reconstructed series of Lamb WTs was used to extend this PC-based NAO index over the period 1781-1996. Although missing data make some early periods unavailable, this index suggests a stronger multidecadal variability of NAO over the 19th century than over the 20th century.

Numerous limitations in our methodology and in our analysis have been underlined in this work. Mainly, the very limited number of daily pressure series provides only weak constraints for reconstructing WTs. Of course this limitation could be improved when homogenised SLP series will become available, or by considering other types of meteorological series (wind direction or speed, for instance). If other series become available over western Europe, and even better over the Atlantic, an objective analysis could be implemented: such an analysis would directly interpret the pressure field as WT, circumventing the problem of calibration (definition of centroids) and of the associated uncertainty. In terms of analysis of our reconstructed series, it was restricted here to the winter season because the climatic impact of WTs is strongest. However, summer variability would also be of interest, especially to study the conditions leading to droughts and heat waves. Finally, we missed

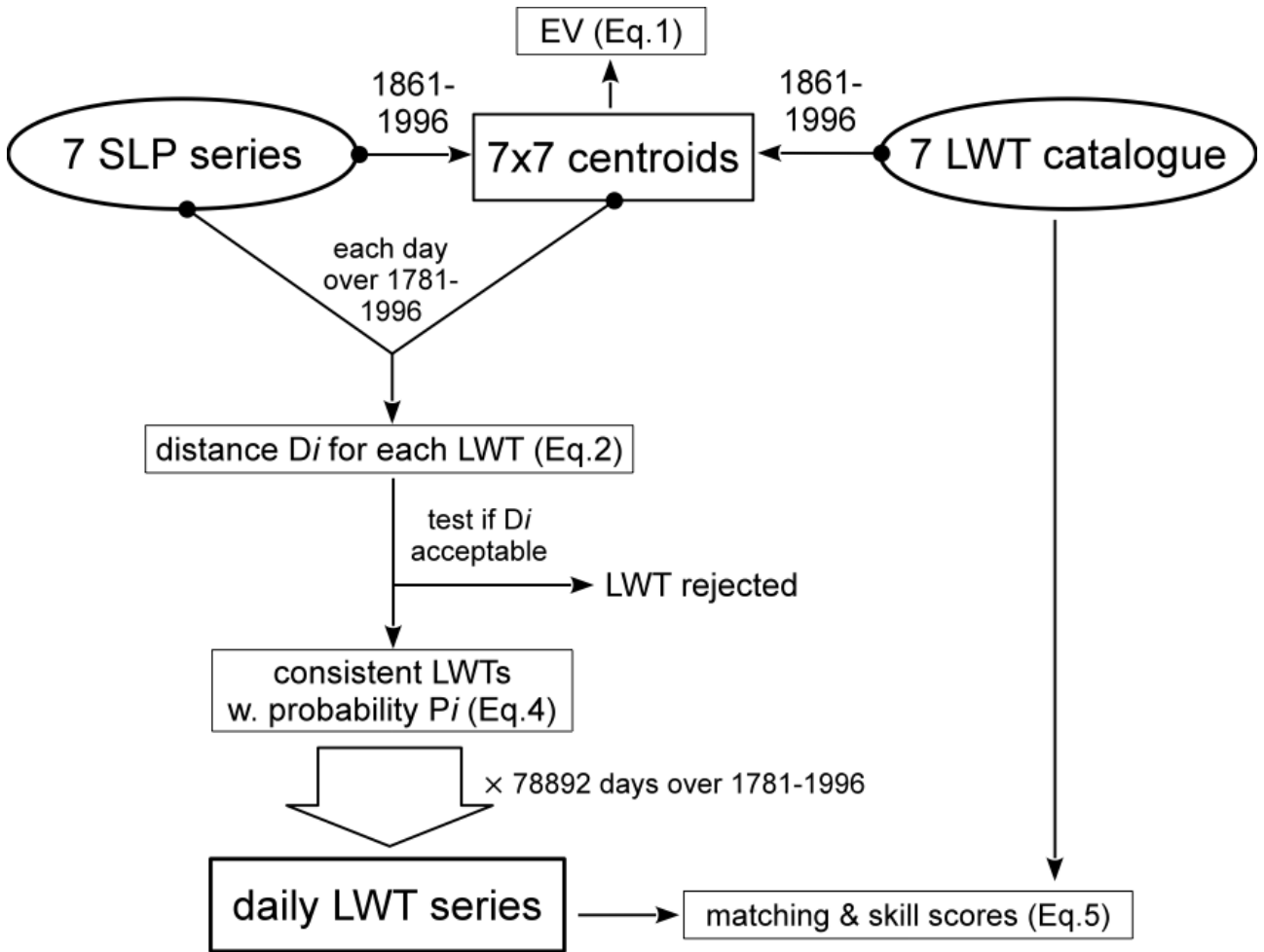
here to address the climatic conditions related to each WT, and the climatic impacts of their variations. Although such quantification is one of the main goals of WTs' work, it is complicated by intra-WT variability and by non-stationarity in their climatic impacts (see, e.g., Hertig et al. 2015).

Acknowledgements We are grateful to meteorologists who spent their careers analysing weather charts and compiling these data; and to 'data savers' who have exhumed and homogenised measurements of surface pressure and made them publicly available. We are also grateful to the reviewers who spent efforts and time to improve the manuscript. MS has been funded by the Swiss SNF project FUPSOL-2; SB acknowledges funding from the SNF project RE-USE.

GD is indebted to the Institute of Geography, University of Bern, for hosting him during a sabbatical stay. Many thanks to the Scilab Consortium for providing the Scilab tool used for calculations and graphs.

Appendix: Details of the classification

1. Flow chart of the classification (example with the objective catalogue)



2. *Calculation of the centroids*: We combine the daily series of pressure measured at the available stations (Table 2) with the daily WTs of the Lamb catalogue ('subjective' catalogue herein), or of the Jones et al. catalogue ('objective' catalogue herein), to calculate the average normalised pressure at each station and for each WT (so-called 'centroids', Fig. 2).

3. *Explained variance (EV)*: The efficiency of the chosen centroids (seven main Lamb WTs) to statistically describe the pressure variability recorded at the stations is quantified with the standard metric of the explained variance (EV):

$$EV = BSS/TSS = 1 - WSS/TSS, \quad (\text{Eq. A1})$$

with BSS the 'between-types-sum-of-squares', WSS the 'within-types-sum-of-squares' and TSS the 'total-sum-of-squares' (e.g., Beck & Philipp, 2010). For these calculations, unclassified days are not

accounted for (accounting for them slightly decreases EV by 1 to 2%). Note that EV depends on the number of WTs considered, hence the values of EV calculated with 7 types are not directly comparable to the ones obtained with 26 types, or with other classification techniques (Beck & Philipp, 2010).

4. *Distance between the daily pressure pattern and centroids*: For each day with pressure data at all stations (otherwise the day is considered as ‘unclassified’), we calculate the Mahalanobis distance D_i between the measured pressure values and the centroids of the corresponding month for the WT i as the sum over the s stations (in matrix form):

$$D_i^2 = (P_s - \bar{P}_i)^T \cdot \text{Cov}_i^{-1} \cdot (P_s - \bar{P}_i) \quad , \quad (\text{Eq. A2})$$

where P_s is the $(s \times 1)$ vector of pressure values measured at the stations for each day, \bar{P}_i is the $(s \times 1)$ vector of centroids (average values) for the weather type i , and Cov_i is the $(s \times s)$ covariance matrix between the pressure values measured at the stations calculated separately for each WT i (as defined in the catalogue).

5. *Probability distribution of the distance*: We use this distance D_i as a metric of the probability distribution of distance to each centroid i . Since the pressure associated to each WT is approximately normally distributed, we assume that the square of the distance D_i^2 follows a chi-squared distribution (e.g., Wilks, 2011), that is,

$$f(D_i^2 / \text{WT} = i) = f\chi_s^2 \quad , \quad (\text{Eq. A3})$$

where $f(D_i^2 / \text{WT} = i)$ is the probability density of D_i^2 for the weather type i , and $f\chi_s^2$ is the probability density of the chi-squared distribution with s degrees of freedom (the number of stations).

6. *Classification consistency*: for each day we evaluate whether the observed pattern of pressure does correctly project onto one of the given seven sets of centroids, and so can be classified. For this, among the seven distances D_i between the observed pressure pattern and the seven WTs, we evaluate whether the shortest distance D_i is short enough to attribute this day to the

corresponding WT i . The formal test is the following. The null hypothesis H_0 to reject is that, for each day, the minimum distance D_i is too large to attribute this day to any WT. Note that by considering the minimum distance for each day, we do not make an *a priori* hypothesis on the matching WT. We accept a risk of wrongly rejecting H_0 of $\alpha = 1\%$. We have to compare the minimum value of the given day with the threshold value corresponding to the 1% of the empirical distribution of the minimum values for all days of the catalogue. For the objective catalogue over the 1861-1996 period, the value which leaves 1% of one-sided, right-tail, probability is 20.1, and is used as the threshold value to reject H_0 . (Note that this empirical distribution follows quite closely a chi-square distribution with 6 degrees of freedom.)

7. *Classification of each day*: if the above test is passed, that is, if there is at least one acceptable WT matching the observed pressure pattern of day j , then we want to calculate the probabilities of the different WTs to match this pressure pattern, P_j , that is, the conditional probabilities of WT i , given P_j : $P(WT_j = i / P_j)$. If day j has to be classified with only one WT, we use the ‘maximum *a posteriori* rule’ to find the WT i which maximizes the probability of the WT i given the pressure pattern P_j .

8. *Joint probabilities of WTs*: for each day j considered as reasonably close to the Lamb WTs, we would like to estimate the probability of each WT i to match the observed pattern of pressure P_j , that is, $P(WT_j = i / P_j)$.

Using Bayes formula and the law of total probability, we write that:

$$P(WT_j = i / P_j) = \frac{f(P_j / WT_j = i) \cdot P(WT_j = i)}{f(P_j)}, \text{ and}$$

$$P(WT_j = i / P_j) = \frac{f(P_j / WT_j = i) \cdot P(WT_j = i)}{\sum_k f(P_j / WT_j = k) \cdot P(WT_j = k)},$$

with k describing all WTs.

The probability of any WT i , $P(WT_j = i)$, is taken as the frequency of this WT over the whole catalogue, F_i .

Last, we assume that the squared distance of the pressure pattern to the centroids follows a chi-square distribution with s degrees of freedom (with s the number of stations), so that:

$$f(P_j / WT_j = i) = f_{\chi^2_s}(D_i^2 / WT_j = i) .$$

Hence we calculate the probabilities $P(WT_j = i / P_j)$ as:

$$P(WT_j = i / P_j) = \frac{f_{\chi^2_s}(D_i^2 / WT_j = i) \cdot F_i}{\sum_k f_{\chi^2_s}(D_k^2 / WT_j = k) \cdot F_k} . \quad (\text{Eq. A4})$$

9. *Skill score of the reconstruction.* The skill score can be calculated as (e.g., Nicholls, 1980):

$$S = \frac{\text{Correct} - \text{Climatology}}{\text{Total} - \text{Climatology}}$$

or in the proportion form:

$$S = \left(\frac{\text{Correct}}{\text{Total}} - \frac{\text{Climatology}}{\text{Total}} \right) / \left(1 - \frac{\text{Climatology}}{\text{Total}} \right) , \quad (\text{Eq. A5})$$

where ‘Correct’ is the number of correctly classified days in our reconstruction with respect to the calibration catalogue, ‘Climatology’ is the number of correctly classified days in a random series of weather types based on their average frequencies, and ‘Total’ the total number of days compared. The skill score is positive if the technique of classification does a better job than classifying days based on their climatology. We estimate the probability that, for any day, the reconstructed weather type WT matches the weather type WTo of the calibration catalogue by chance as:

$$P(WT = WTo) = \sum_{i=1}^{n+1} P(WT = i) \cdot P(WTo = i) , \quad (\text{Eq. A6})$$

where i is one of the n WTs (plus the unclassified type). We estimate the probability of any WT i , $P(WT=i)$ or $P(WTo=i)$, as its average frequency in the reconstructed series or in the calibration catalogue, respectively (Table 3). Since we also calculated the matching score by considering the most likely two WTs, WT_1 and WT_2 , we write the combined probability by chance as:

$$P(WT_1 = WTo \cup WT_2 = WTo) = \sum_{i=1}^{n+1} (P(WT_1 = i) \cdot P(WTo = i) + P(WT_2 = i) \cdot P(WTo = i))$$

and because WT_1 and WT_2 are different, this reads:

$$P(WT_1=WT_o \cup WT_2=WT_o) = 2 \times \sum_{i=1}^{n+1} P(WT_1=i) \cdot P(WT_o=i) \quad (\text{Eq. A7})$$

This ‘climatology’ probability is found to be 18% for one WT (i.e., $P(WT=WT_o)$), and 36% for two WTs (i.e., $P(WT_1=WT_o \text{ OR } WT_2=WT_o)$), throughout the different reconstructions.

References

- Ansell TJ et al (2006) Daily Mean Sea Level Pressure Reconstructions for the European-North Atlantic Region for the Period 1850-2003. *J.Climate* 19:2717-2742. doi:10.1175/JCLI3775.1
- Attmannspacher W (1981) 200 Jahre meteorologische Beobachtungen auf dem Hohenpeißenberg 1781-1980. *Berichte Deutscher Wetterdienst* 155, Offenbach am Main, 222p.
- Barriopedro D et al (2014) Witnessing North Atlantic westerlies variability from ships' logbooks (1685-2008). *Clim Dyn* 43:939-955. doi:10.1007/s00382-013-1957-8
- Beck C, Jacobeit J, Jones PD (2007) Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central europe since 1780. *Int J Climatol* 27:473-491. doi:10.1002/joc.1410
- Beck C, Philipp A (2010) Evaluation and comparison of circulation type classifications for the European domain. *Phys Chemis Earth* 35:374-387. doi:10.1016/j.pce.2010.01.001
- Bergström H, Moberg A (2002) Daily Air Temperature and Pressure Series for Uppsala (1722-1998). *Climatic Change* 53:213-252. doi:10.1023/A:1014983229213
- Branstator G, Selten F (2009) "Modes of Variability" and Climate Change. *J Climate* 22:2639-2658. doi:10.1175/2008JCLI2517.1
- Brönnimann S et al (2013) Historical weather extremes in the "Twentieth Century Reanalysis". In: Brönnimann S and Martius O (ed) *Weather extremes during the past 140 years*, Geographica Bernensia, Bern, CH, pp 7-17
- Brugnara Y et al (2015) A collection of sub-daily pressure and temperature observations for the early instrumental period with a focus on the "year without a summer" 1816. *Clim Past* 11, 1027-1047. doi:10.5194/cp-11-1027-2015
- Compo GP et al (2011) The Twentieth Century Reanalysis Project. *Q J R Meteorol Soc* 137:1-28. doi:10.1002/qj.776
- Cornes RC et al (2012a) A daily series of mean sea-level pressure for London, 1692-2007. *Int J Climatol* 32:641-656. doi:10.1002/joc.2301

- Cornes RC et al (2012b) A daily series of mean sea-level pressure for Paris, 1670-2007. *Int J Climatol* 32:1135-1150. doi:10.1002/joc.2349
- Cornes RC et al (2013) Estimates of the North Atlantic Oscillation back to 1692 using a Paris-London westerly index. *Int J Climatol* 33:228-248. doi:10.1002/joc.3416
- Corti S, Molteni F, Palmer TN (1999) Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature* 398:799-802. doi:10.1038/19745
- D'Arrigo R et al (2011) The anomalous winter of 1783-1784: Was the Laki eruption or an analog of the 2009-2010 winter to blame? *Geophys Res Lett* 38:L05706. doi:10.1029/2011GL046696
- Ferguson CR, Villarini G (2013) An evaluation of the statistical homogeneity of the Twentieth Century Reanalysis. *Climate Dynamics* 42:2841-2866. doi:10.1007/s00382-013-1996-1
- Goodkin NF et al (2008) Increased multidecadal variability of the North Atlantic Oscillation since 1781. *Nature Geosci* 1:844-848. doi:10.1038/ngeo352
- Hertig E et al (2015) A review of non-stationarities in climate variability of the last century with focus on the North Atlantic - European sector. *Earth-Sci Rev* 147:1-17. doi:10.1016/j.earscirev.2015.04.009
- Hess P, Brezowsky H (1977) Katalog der Grosswetterlagen Europas 1881-1976. 3. verbesserte und ergänzte Auflage. Deutscher Wetterdienst, Offenbach am Main, Germany
- Hurrell JW (1995) Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science* 269:676-679. doi:10.1126/science.269.5224.676
- Jenkinson AF, Collison FP (1977) An initial climatology of gales over the North Sea. U.K. Meteorological Office, Bracknell, UK
- Jones PD, Wigley TML, Briffa KR (1987) Monthly mean pressure reconstructions for Europe (back to 1780) and North America (to 1858). U.S. Dept. of Energy, Office of Energy Research, Carbon Dioxide Research Division, Washington, DC
- Jones PD, Hulme M, Briffa KR (1993) A comparison of Lamb circulation types with an objective classification scheme. *Int J Climatol* 13:655-663. doi:10.1002/joc.3370130606

- Jones PD, Osborn TJ, Briffa KR (2003) Pressure-Based Measures of the North Atlantic Oscillation (NAO): A Comparison and an Assessment of Changes in the Strength of the NAO and in its Influence on Surface Climate Parameters. In: Hurrell JW, et al (ed) *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, American Geophysical Union, Washington, D.C., pp 51-62
- Jones PD, Harpham C, Briffa KR (2013) Lamb weather types derived from reanalysis products. *Int J Climatol* 33:1129-1139. doi:10.1002/joc.3498
- Jones PD et al (2014) The development of Lamb weather types: from subjective analysis of weather charts to objective approaches using reanalyses. *Weather* 69:128-132. doi:10.1002/wea.2255
- Kington JA (1970) A late eighteenth century source of meteorological data. *Weather* 25:169-175. doi:10.1002/j.1477-8696.1970.tb05224.x
- Kington JA (1988) *The Weather of the 1780s over Europe*. Cambridge University Press, Cambridge, UK
- Krueger O et al (2013) Inconsistencies between Long-Term Trends in Storminess Derived from the 20CR Reanalysis and Observations. *J Climate* 26:868-874. doi:10.1175/JCLI-D-12-00309.1
- Lamb HH (1950) Types and spells of weather around the year in the British Isles : Annual trends, seasonal structure of the year, singularities. *Q J R Meteorol Soc* 76:393-429. doi:10.1002/qj.49707633005
- Lamb HH (1972) *British Isles weather types and a register of daily sequence of circulation patterns 1861-1971*. H.M. Stationery Office, London, UK.
- Lamb HH (1994) *British Isles Daily Wind and Weather Patterns 1588, 1781–86, 1991 and shorter early sequences*. *Clim Monitor* 20:47-71.
- Lamb HH, Johnson AI (1966) *Secular variations of the atmospheric circulation since 1750*. Meteorological Office, HMSO, London, UK.
- Maugeri M et al (2002) Daily Milan Temperature and Pressure Series (1763-1998): Completing and Homogenising the Data. *Clim Change* 53:119-149. doi:10.1023/A:1014923027396

- Michelangeli P-A, Vautard R, Legras B (1995) Weather Regimes: Recurrence and Quasi Stationarity. *J Atmos Sci* 52:1237-1256.
- Moberg A et al (2002) Daily Air Temperature and Pressure Series for Stockholm (1756-1998). *Climatic Change* 53:171-212. doi:10.1023/A:1014966724670
- Murray R, Lewis RPW (1966) Some aspects of the synoptic climatology of the British Isles as measured by simple indices. *Meteorol Mag* 95:193- 203.
- Nicholls N (1980) Long-range weather forecasting: Value, status, and prospects. *Rev Geophys* 18:771-788. doi:10.1029/RG018i004p00771
- Ortega P et al (2015) A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature* 523:71-74. doi:10.1038/nature14518
- Osborn TJ et al (1999) Evaluation of the North Atlantic Oscillation as simulated by a coupled climate model. *Clim Dyn* 15:685-702. doi:10.1007/s003820050310
- Perry A (1969) The P.S.C.M. Index and Regional Anomalies of Temperature, Rainfall and Sunshine. *Weather* 24:225-228. doi:10.1002/j.1477-8696.1969.tb03184.x
- Schwander M et al (2017) Reconstruction of Central European daily weather types back to 1763. *Int J Climatol* 37:30-44. doi:10.1002/joc.4974
- Slonosky VC, Jones PD, Davies TD (2000) Variability of the surface atmospheric circulation over Europe, 1774-1995. *Int J Climatol* 20:1875-1897. doi:10.1002/1097-0088(200012)20:15<1875::AID-JOC593>3.0.CO;2-D
- Sweeney JC, O'Hare GP (1992) Geographical Variations in Precipitation Yields and Circulation Types in Britain and Ireland. *Trans Inst Br Geogr* 17:448-463. doi:10.2307/622710
- Vinther BM et al (2003) Improving the Gibraltar/Reykjavik NAO index. *Geophys Res Lett* 30:2222. doi:10.1029/2003GL018220
- Wallace JM (2000) North atlantic oscillation / annular mode: Two paradigms - one phenomenon. *Q J R.Meteorol Soc* 126:791-805. doi:10.1002/qj.49712656402
- Wanner H et al (2001) North Atlantic Oscillation – Concepts And Studies. *Surv Geophys* 22:321-

381. doi:10.1023/A:1014217317898

Werner PC et al (2000) Recent climate change in the North Atlantic/European sector. *Int J Climatol* 20:463-471. doi:10.1002/(SICI)1097-0088(200004)20:5<463::AID-JOC483>3.0.CO;2-T

Werner PC, Gerstengarbe F-W (2011) Spatial-temporal changes of meteorological parameters in selected circulation patterns. Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

Wilby RL (1993) Evidence of ENSO in the synoptic climate of the British Isles since 1880. *Weather* 48:234-239. doi:10.1002/j.1477-8696.1993.tb05897.x

Wilby R (1995) Simulation of precipitation by weather pattern and frontal analysis. *J Hydrol* 173:91-109. doi:10.1016/0022-1694(95)02715-2

Wilby RL, O'Hare G, Barnsley N (1997) The North Atlantic Oscillation and British Isles climate variability, 1865-1996. *Weather* 52:266-276. doi:10.1002/j.1477-8696.1997.tb06323.x

Wilby RL, Quinn NW (2013) Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. *J Hydrol* 487:109-121. doi:10.1016/j.jhydrol.2013.02.038

Wilks DS (2011) Chapter 11 - The Multivariate Normal (MVN) Distribution. In: Wilks DS (ed) *Statistical Methods in the Atmospheric Sciences*, Academic Press, pp 491-518

Woollings T (2010) Dynamical influences on European climate: an uncertain future. *Phil Trans R Soc A* 368:3733-3756. doi:10.1098/rsta.2010.0040

Yiou P et al (2007) Inconsistency between atmospheric dynamics and temperatures during the exceptional 2006/2007 fall/winter and recent warming in Europe. *Geophys Res Lett* 34:L21808. doi:10.1029/2007GL031981

Lamb's 26 weather types	Frequency of the 26 types in the subjective catalogue (%)	Corresponding 7 main types	Frequency of the 7 types in the subjective catalogue (%)	Frequency of the 7 types in the objective catalogue (%)
Anticyclonic (A)	18.3	Anticyclonic (A)	33.0	33.9
Anticyc. North-easterly (ANE)	1.4			
Anticyc. Easterly (AE)	2.4			
Anticyc. South-easterly (ASE)	1.0			
Anticyc. Southerly (AS)	1.1			
Anticyc. South-westerly (ASW)	0.9			
Anticyc. Westerly (AW)	4.6			
Anticyc. North-westerly (ANW)	1.5			
Anticyc. Northerly (AN)	2.0			
North-easterly (NE)	1.0	<i>Unclassified</i>		
Easterly (E)	3.6	Easterly (E)	3.6	4.0
South-easterly (SE)	1.7	<i>Unclassified</i>		
Southerly (S)	4.3	Southerly (S)	4.3	11.7
South-westerly (SW)	2.9	<i>Unclassified</i>		
Westerly (W)	18.3	Westerly (W)	18.3	14.1
North-westerly (NW)	3.7	North-westerly (NW)	3.7	5.0
Northerly (N)	4.6	Northerly (N)	4.6	4.2
Cyclonic (C)	13.1	Cyclonic (C)	23.0	25.6
Cyclonic North-easterly (CNE)	0.4			
Cyclonic Easterly (CE)	1.1			
Cyclonic South-easterly (CSE)	0.5			
Cyclonic Southerly (CS)	1.2			
Cyclonic South-westerly (CSW)	0.7			
Cyclonic Westerly (CW)	3.9			
Cyclonic North-westerly (CNW)	0.9			
Cyclonic Northerly (CN)	1.3			
<i>Unclassified</i>	3.9	<i>Unclassified</i>	9.5	1.6

Table 1. The 26 weather types of Lamb's catalogue, and the derived 7 main weather types using the simplification scheme favouring the A&C types. Frequencies are calculated over the period 1.1.1861 to 3.2.1997 for the Lamb catalogues, or over the period 1.1.1871 to 10.4.2015 for the objective catalogue.

Station	Daily parameter used in this study	Period of coverage	missing days over 1861-1996 / 1781-1996	Reference
London (UK)	SLP	1.1.1692 - 31.12.2007	13 / 27	Cornes et al., 2012a
Paris (Fr)	SLP	1.1.1670 - 31.12.2007	161 / 262	Cornes et al., 2012b
Stockholm (Sw)	SLP	1.1.1756 - 31.12.2000	0 / 0	Moberg et al., 2002
Uppsala (Sw)	SLP	1.12.1722 - 31.12.2000	0 / 0	Bergström & Moberg, 2002
Hohenpeißenberg (Ge)	SLP	1.1.1781 - 31.12.2007	11 / 2221	Deutscher Wetterdienst / Attmannspacher 1981
Amsterdam (Ne)	SLP	1.1.1743 - 31.10.2014	0 / 298	Brandsma et al., pers. com. 2015
Milan (It)	SLP	1.1.1763 - 31.12.1998	0 / 0	Maugeri et al., 2002

Table 2. List of daily sea level pressure (SLP) records used in this study. ‘Missing days’ is the number of days without data in each original record.

Main 7 weather types	subjective catalogue	reconstructed series	correctly classified days	objective catalogue	reconstructed series	correctly classified days
North-westerly	3.8	7.7	3.3	4.8	8.8	4.4
Westerly	18.3	14.5	15.8	13.9	10.3	10.1
Southerly	4.3	4.5	3.4	12.0	8.6	8.4
Easterly	3.6	7.3	4.1	4.1	6.6	3.9
Northerly	4.6	11.3	5.7	4.3	11.4	4.8
Cyclonic	23.0	24.4	28.1	25.4	22.4	29.9
Anticyclonic	33.0	23.9	39.4	34.0	25.5	38.5
<i>Unclassified</i>	9.5	6.3	0.2	1.6	6.4	0.1

Table 3. Frequencies (in %) of WTs in the catalogue used for calibration, in the total reconstructed series, and in the correctly classified days. Period of calibration and calculation: 1861-1996 for the subjective catalogue; 1871-1996 for the objective catalogue. (Note that the values for the catalogues are slightly different from the ones in Table 1 since they are calculated over different periods)

metric:	matching score		skill score S	
	most likely WT	most likely two WTs	most likely WT	most likely two WTs
number of WT tested per day:				
Subjective catalogue (1861-1996), calibration over 1861-1996	56%	75%	0.47	0.61
Subjective catalogue (1861-1996), calibration over 1945-1996	55%	74%	0.44	0.58
Subjective catalogue 1861-1944, calibration over 1945-1996	52%	72%	0.41	0.54
Objective catalogue (1871-1996), calibration over 1871-1996	59%	78%	0.50	0.65
Objective catalogue (1871-1996), calibration over 1945-1996	57%	76%	0.48	0.62
Objective catalogue 1871-1944, calibration over 1945-1996	54%	72%	0.43	0.56
Additional periods analysed by Lamb et al. (1951 days with 7 WTs)	46%	64%	(0.39)	(0.53)

Table 4. Evaluation of the reconstructed WT series with respect to the calibration catalogue (i.e., used to calculate the centroids over the indicated period). Last row: WTs calibrated to the subjective catalogue and compared to short classified periods published by Lamb (1994). Parentheses for skill score values emphasize that these values are also affected by the difference of WT frequencies between these short periods and the other calibration periods. See the Appendix for the calculation of the matching and skill scores.

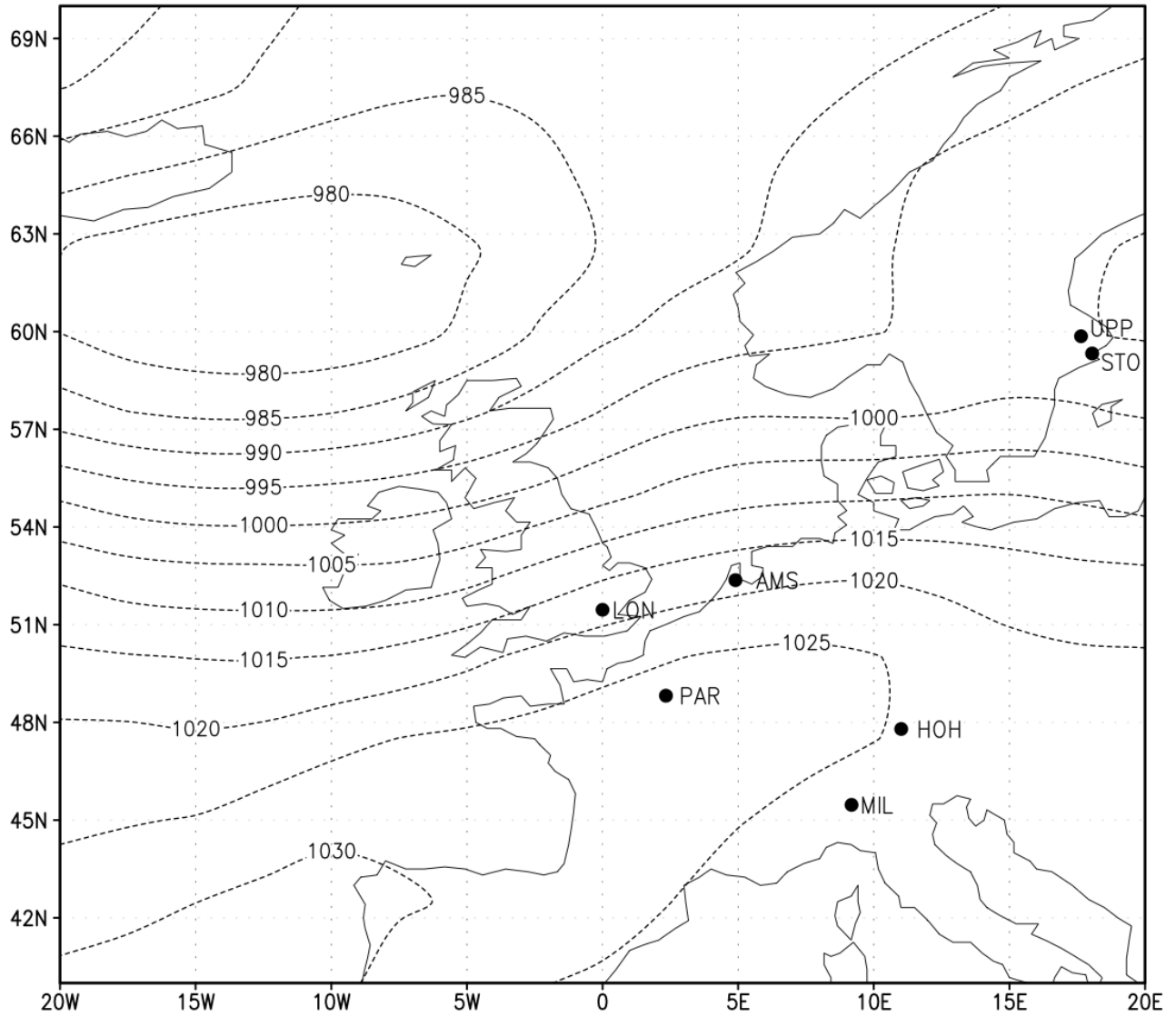


Fig. 1 Sea level pressure for the 22nd February 1995 (NCEP reanalysis II), a situation classified by Hubert Lamb as ‘Westerly flow’ (W). Dots locate the stations used in this study (Table 2), labelled as: AMS=Amsterdam, HOH=Hohenpeißenberg, MIL=Milan, LON=London, PAR=PARIS, STO=Stockholm, UPP=Uppsala.

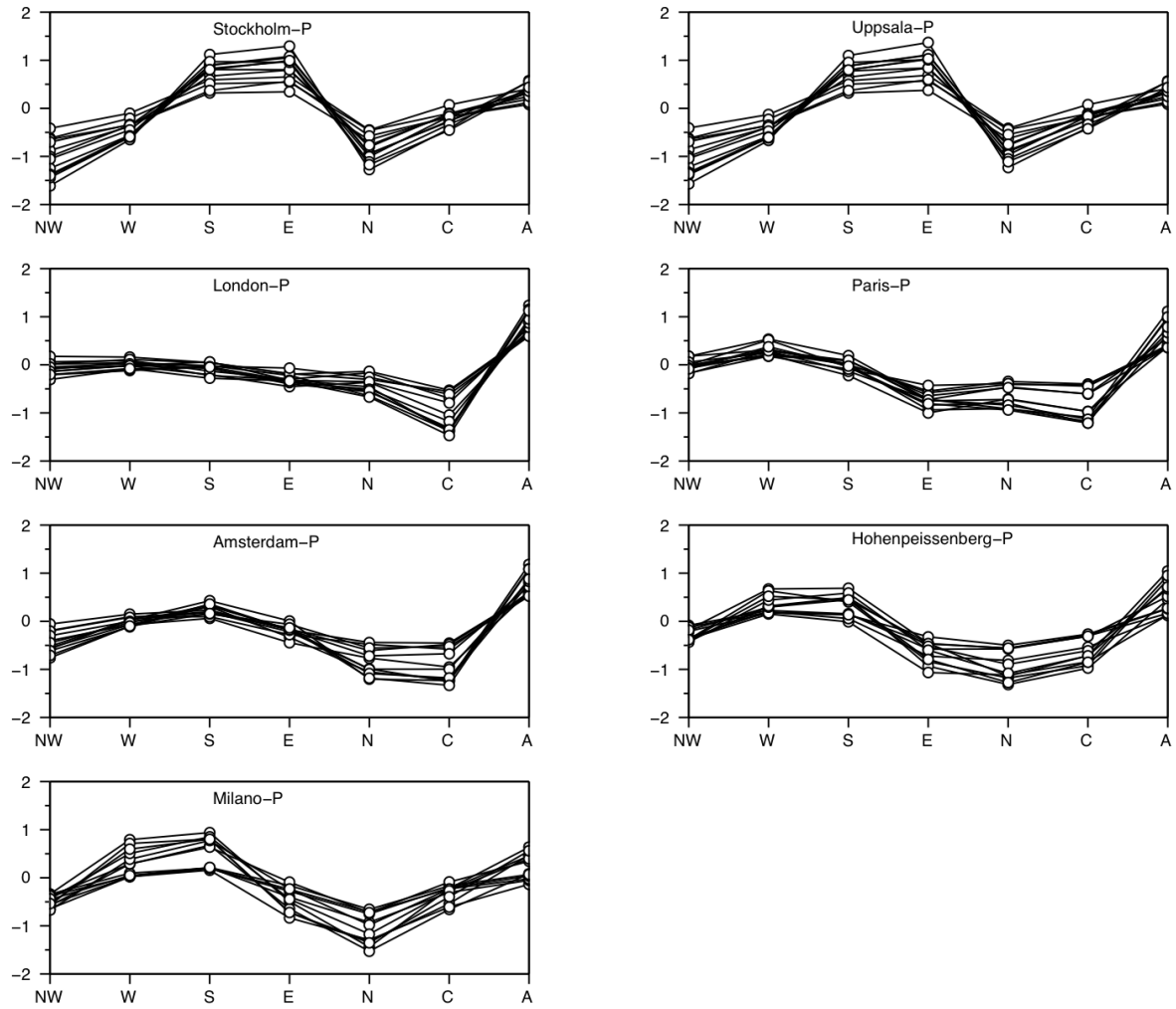


Fig. 2 Centroids of the normalised pressure at each station, for each month and each of the seven LWT of the subjective catalogue, averaged over the period 1861-1996 (see Table 1 for the WT codes).

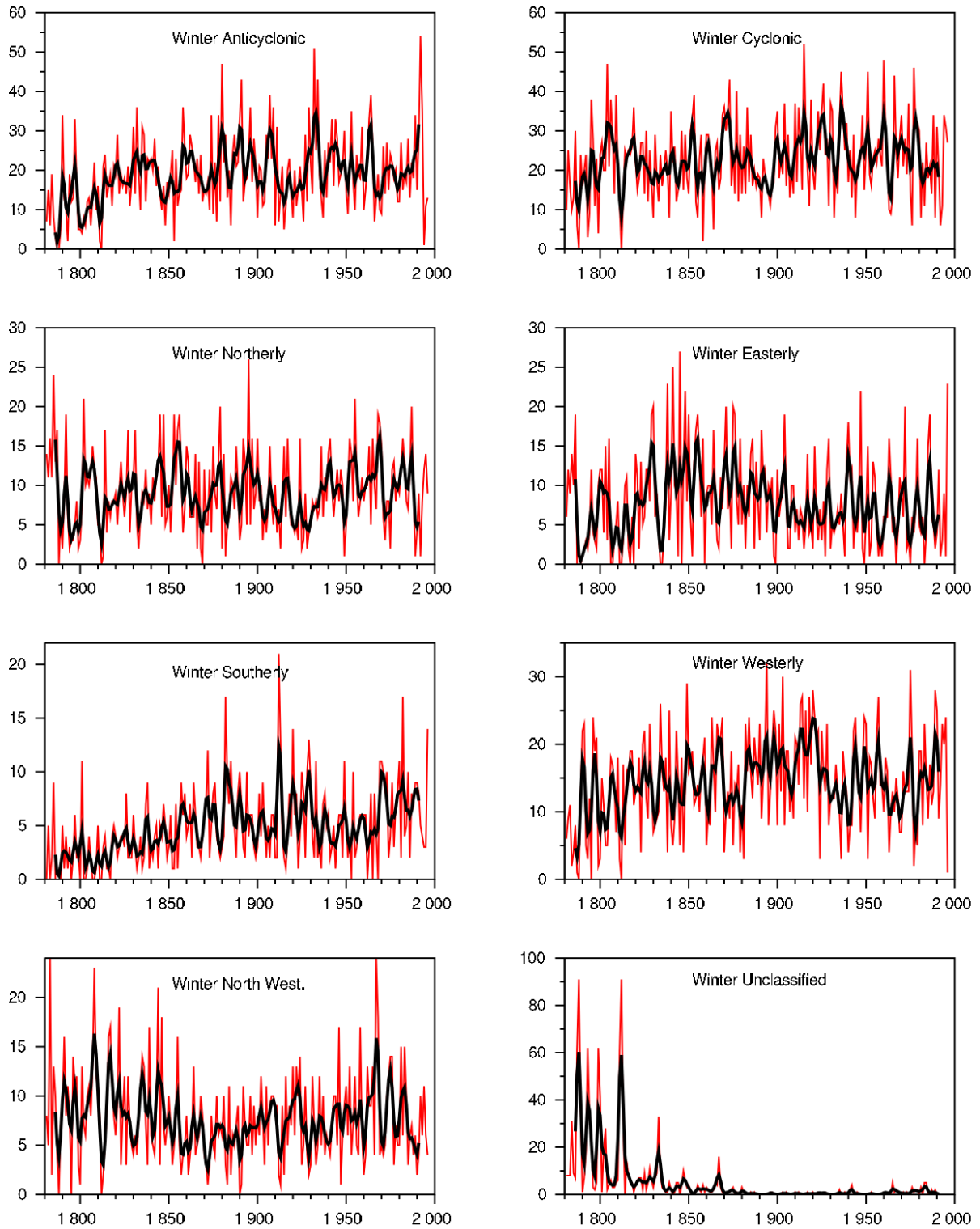


Fig. 3 The number of days per winter (DJF, 90 days) with each Lamb WT in the reconstructed series based on the subjective catalogue, over the period 1781-1996. The black curve is an 11-yr Gaussian average. Note that, for readability, each graph has its own scale.

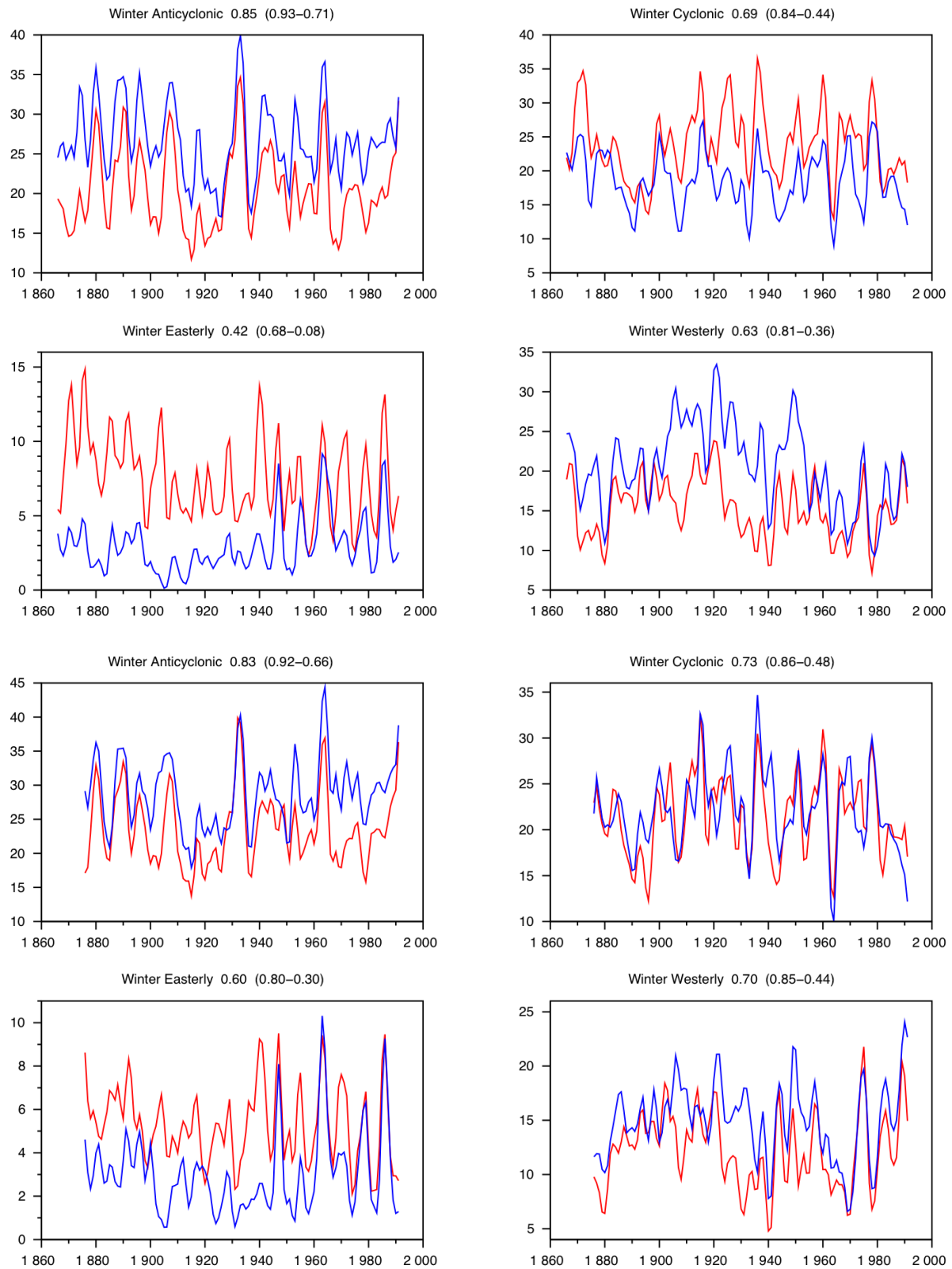


Fig. 4 Number of days per winter (DJF) with four weather types, in the subjective (top) or objective (bottom) catalogue (in blue), compared to our reconstructed series (in red), smoothed with a 11-yr Gaussian filter. Correlation coefficients (r) between the smoothed series are indicated with the Fisher 95% confidence interval accounting for the effective numbers of degrees of freedom.

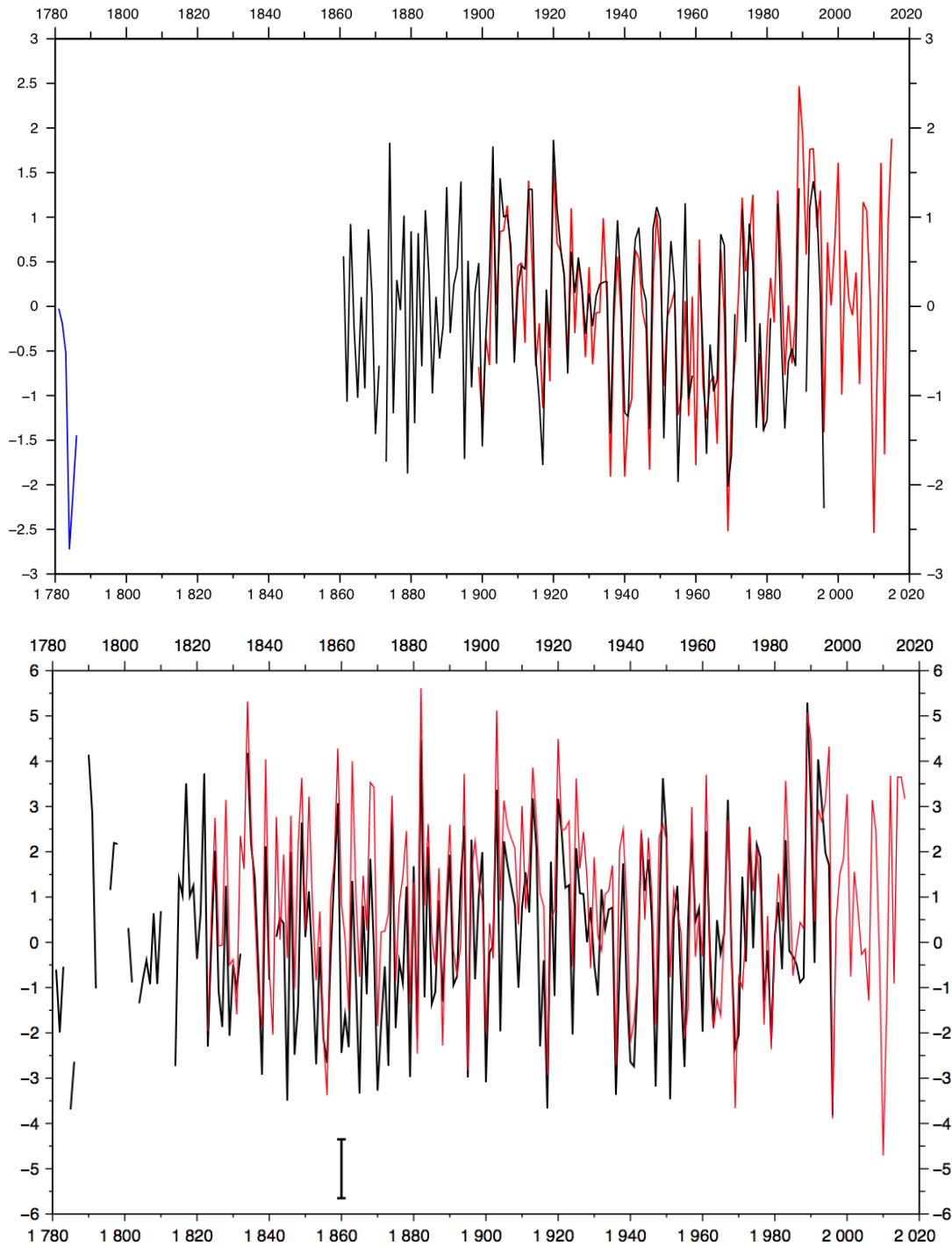


Fig. 5 Time series of winter (DJFM) North Atlantic Oscillation indices. Top (a): In red, the PC-based index of Hurrell (1995). In black, our index based on WT frequencies from the subjective Lamb catalogue over 1781-1786 and 1861-1996 regressed to the PC-based NAO index. Bottom (b): In red, the Iberia-Reykjavik index of Vinther et al. (2003). In black, our index based on Lamb weather type frequencies reconstructed from stations records over the 1781-1996 period. No value plotted for DJFM periods with more than 24 unclassified days (i.e., 20% of the period). The error bar is the ± 1 sigma uncertainty on our index. Both indices have been normalised to the index it is compared with for reading purpose (top over 1899-1996 with PC-based index, and bottom over 1951-1980 with the Vinther et al. index).